CLAIMS

What is claimed is:

1. A computer-implemented method for optimizing allocation of computer resources, comprising:

establishing a server model including one or more server nodes, wherein each server node has an associated set of capacity attributes;

establishing a service model including one or more service nodes, wherein each service node has an associated set of demand attributes;

selecting one of a plurality of user-selectable optimization methods; and

generating an optimized mapping of the server nodes in the server model to the service nodes in the service model using the selected optimization method and demand and capacity attributes.

2. The method of claim 1, further comprising normalizing the capacity attributes of server nodes and demand attributes of service nodes.

3. The method of claim 2, further comprising:

selecting one of a plurality of user-selectable objective functions, wherein each of the objective functions evaluates a solution as a function of the demand and capacity attributes; and

generating an optimized mapping of the server nodes to the service nodes using the selected one of the objective functions and selected one of the optimization methods.

4. The method of claim 3, wherein the plurality of optimization methods include a genetic process and a complete search process.

5. The method of claim 3, further comprising:

establishing one or more service-node relationships between selected pairs of the service nodes, wherein each service-node relationship has an associated transport demand attribute specifying a quantity of communication resources required for communication between the associated pair of service nodes;

establishing one or more server-node relationships between selected pairs of the server nodes, wherein each server-node relationship has an associated transport capacity attribute specifying a quantity of communication resources available for communication between the associated pair of server nodes; and

generating the optimized mapping as a function of the service-node relationships and server-node relationships.

6. The method of claim 4, wherein the plurality of objective functions includes a first function for quantifying a balance processing load between nodes and a second function quantifies a transport demand between the nodes.

7. The method of claim 4, further comprising:

establishing a plurality of server models, each server model including one or more server nodes, wherein each server node has an associated set of capacity attributes;

designating a layered relationship between the server models, wherein for a first server-model layer immediately above a second server-model layer, the second server-model layer includes respective models that represent the nodes in the first server-model layer;

establishing a plurality of service models, each service model including one or more service nodes, wherein each service node has an associated set of demand attributes; and

designating a layered relationship between the service models, wherein for a first service-model layer immediately above a second service-model layer, the second service-model layer includes respective models that represent the nodes in the first server-model layer;

normalizing the capacity attributes of server nodes and demand attributes of service nodes of the server models and the service models, respectively; and

generating an optimized mapping of the server nodes in a user-selected one of the server models to service nodes in a user-selected one of the service models using the selected optimization method.

8. The method of claim 7, further comprising:

establishing one or more service-node relationships between selected pairs of the service nodes, wherein each service-node relationship has an associated transport demand attribute specifying a quantity of communication resources required for communication between the associated pair of service nodes;

establishing one or more server-node relationships between selected pairs of the server nodes, wherein each server-node relationship has an associated transport capacity attribute specifying a quantity of communication resources available for communication between the associated pair of server nodes; and

generating the optimized mapping as a function of the service-node relationships and server-node relationships.

9. The method of claim 7, wherein each service node has an associated set of capacity attributes and further comprising generating an optimized mapping of service nodes in a first user-selected service model to service nodes in a second user-selected service model as a function of the demand attributes of the first service model and capacity attributes of the second service model.

10. The method of claim 7, wherein each server node has an associated set of demand attributes and further comprising generating an optimized mapping of server nodes in a first user-selected server model to server nodes in a second user-selected server model as a function of the demand attributes of the first server model and capacity attributes of the second server model.

11. The method of claim 1, further comprising:

selecting one of a plurality of user-selectable objective functions, wherein each of the objective functions evaluates a solution as a function of the demand and capacity attributes; and

generating an optimized mapping of the server nodes to the service nodes using the selected one of the objective functions and selected one of the optimization methods.

12. The method of claim 1, wherein the plurality of optimization methods includes a genetic process and a complete search process.

13. The method of claim 1, further comprising:

establishing a plurality of server models, each server model including one or more server nodes, wherein each server node has an associated set of capacity attributes;

designating a layered relationship between the server models, wherein for a first server-model layer immediately above a second server-model layer, the second server-model layer includes respective models that represent the nodes in the first server-model layer;

establishing a plurality of service models, each service model including one or more service nodes, wherein each service node has an associated set of demand attributes; and

designating a layered relationship between the service models, wherein for a first service-model layer immediately above a second service-model layer, the second service-model layer includes respective models that represent the nodes in the first server-model layer;

normalizing the capacity attributes of server nodes and demand attributes of service nodes of the server models and the service models, respectively; and

generating an optimized mapping of the server nodes in a user-selected one of the server models to service nodes in a user-selected one of the service models using the selected optimization method.

14. An apparatus for optimizing allocation of computer resources, comprising:

means for establishing a server model including one or more server nodes, wherein each server node has an associated set of capacity attributes;

means for establishing a service model including one or more service nodes, wherein each service node has an associated set of demand attributes;

means for selecting one of a plurality of user-selectable optimization methods; and

means for generating an optimized mapping of the server nodes in the server model to the service nodes in the service model using the selected optimization method and demand and capacity attributes.

15. A system for identifying optimal allocations of computing resources in a data processing arrangement having a plurality of computing machines that host a plurality of application processes, comprising:

a model repository including a plurality of server models and a plurality of service models, each server model including one or more server nodes and each server node having an associated set of normalized capacity attributes, each service model including one or more service nodes and each service node having an associated set of normalized demand attributes, wherein the server models are defined in a layered relationship and for a first server-model layer immediately above a second server-model layer, the second server-model layer includes respective models that represent the nodes in the first server-model layer, and the service models are defined in a layered relationship and for a first service-model layer immediately above a second service-model layer, the second service-model layer includes respective models that represent the nodes in the first service-model layer; and

an optimization engine coupled to the model repository, the optimization engine including a plurality of user-selectable objective functions and a plurality of user-selectable optimization methods, wherein each of the objective functions evaluates a mapping as a function of the demand and capacity attributes, and each of the optimization methods generates mappings of service nodes in a user-selected service model to server nodes in a user-selected server model and selects an optimal one of the mappings.

16. The system of claim 15, further comprising:

wherein the model repository further includes one or more service-node relationships between selected pairs of the service nodes, each service-node relationship having an associated transport demand attribute that specifies a quantity of communication resources required for communication between the associated pair of service nodes;

wherein the model repository further includes one or more server-node relationships between selected pairs of the server nodes, each server-node relationship having an associated transport capacity attribute that specifies a quantity of communication resources available for communication between the associated pair of server nodes; and

the optimization engine is further configured to generate the optimized mapping as a function of the service-node relationships and server-node relationships.

17. The system of claim 15, wherein each service node has an associated set of capacity attributes and the optimization engine is further configured to generate an optimized mapping of service nodes in a first user-selected service model to service nodes in a second user-selected service model as a function of the demand attributes of the first service model and capacity attributes of the second service model.

18. The system of claim 15, wherein each server node has an associated set of demand attributes and the optimization engine is further configured to generate an optimized mapping of server nodes in a first user-selected server model to server nodes in a second user-selected server model as a function of the demand attributes of the first server model and capacity attributes of the second server model.